Breast Cancer Screening Services:
Trade-offs in Quality, Capacity, Outreach, and Centralization

Short Title: Breast Cancer Screening

Evrim D. Güneş, Stephen E. Chick, INSEAD, Fontainebleau, France

and O. Zeynep Akşin, Koç University, İstanbul, Turkey

Corresponding author: Evrim D. Güneş

Before Sep 1, 2004

INSEAD

Technology Management Area,

Boulevard de Constance

77305 Fontainebleau CEDEX, France

evrim-didem.gunes@insead.edu

Tel:+(33) 1.60.72.40.46

Fax:+(33) 1.60.74.55.00

After Sep 1, 2004

Operations Management Group

College of Administrative Sciences and Economics

Koç University

Rumeli Feneri Yolu

80910 Sariyer-İstanbul-Turkey

Tel: (90-212) 338 15 45

Breast Cancer Screening Services:
Trade-offs in Quality, Capacity, Outreach, and Centralization

Short Title: Breast Cancer Screening

Abstract

This work combines and extends previous work on breast cancer screening models by explicitly incorporating, for the first time, aspects of the dynamics of health care states, program outreach, and the screening volume-quality relationship in a service system model to examine the effect of public health policy and service capacity decisions on public health outcomes. We consider the impact of increasing standards for minimum reading volume to improve quality, expanding outreach with or without decentralization of service facilities, and the potential of queueing due to stochastic effects and limited capacity. The results indicate a strong relation between screening quality and the cost of screening and treatment, and emphasize the importance of accounting for service dynamics when assessing the performance of health care interventions. For breast cancer screening, increasing outreach without improving quality and maintaining capacity results in less benefit than predicted by standard models.

Keywords: breast cancer screening; system dynamics model; volume-quality relationship; queueing; public health policy; public health outcomes; mammogram; health care services.

1 Introduction

Breast cancer is the most common cancer among women, and the second leading cause of cancer related deaths after lung cancer. The World Health Organization (WHO) estimates that more than 1.2 million people were diagnosed with breast cancer worldwide in 2001 [1]. The American Cancer Society (ACS) estimated that breast cancer would be diagnosed in 211,300 women, and 39,800 would die from the disease in 2003 in the U.S. [2]. The value of mass screening and the ages at which it is appropriate are currently in dispute [3]. However most developed countries have organized screening programs to pro-actively detect breast cancer [4], as early diagnosis of most types of breast cancer is very effective. The 5 year survival rate is 98% with

early stage breast cancer treatment [2]. This paper contributes to the understanding of factors that influence breast cancer screening effectiveness.

While many have examined breast cancer screening in operations [5-8], statistics [9-12], and health [4, 13-17], this paper appears to be the first to include aspects of all of the following interacting effects in the same model: (i) disease progression, (ii) the link between service quality and the volume of mammogram screens provided by the health care provider, (iii) participation levels in the health care program, (iv) factors influencing the participation in a mammogram screening program, and (v) limited service capacity and the potential effect of the utilization of service resources on patient waiting and an increased potential for poor health outcomes due to late diagnosis. Accounting for system dynamics can strongly influence model-based health policy decisions ([18, 19], others below), so we do so here. The first of four model-based experiments in this paper assesses the cost implications for two approaches to improving early detection: outreach and quality improvements. The second studies interactions between participation levels and the potentially deleterious health effects of waiting due to stochastic effects and highly utilized capacity. The third and fourth examine interactions between service decentralization, access, and screening quality. The analysis shows that increasing outreach without improving quality and maintaining capacity may result in less beneficial results than predicted by standard models due to the interactions of these effects.

Critical factors for breast cancer screening program success include two quality measures, *sensitivity* (the probability of detecting cancer in a patient with the disease) and *specificity* (the probability of a negative result in a patient without the disease), as well as *acceptability*, the extent to which those for whom the test is designed have access to and agree to participate in testing [20]. Figure 1 summarizes some interactions between these factors: quality, access, system capacity, and health outcomes. *Screening quality* is influenced by radiologist experience, the annual volume of readings, and film quality standards, among other factors. The popular press [21] recently highlighted problems with screening quality in the U.S. and the importance of the experience of radiologists who read mammograms. One potential cause identified is low minimum

accreditation standards: 480 mammograms readings per year [22] compared with 3,000/year in British Columbia, Canada and 5,000/year in the U.K. [23]. Some argue that radiologists should read a minimum of 2,500 mammograms per year to stay sharp [23]. While imperfect reading quality is perhaps inevitable, the lack of quality incurs system costs. False positives add extra cost and consume service capacity for follow-up tests, incur the potential for unnecessary treatment, and can burden patients [24]. False negative results decrease the chances of survival by missing the opportunity for early detection and treatment.

FIGURE 1 NEAR HERE

For *acceptability/access*, the WHO [20] recommends that mammography should not be introduced for breast cancer screening unless the resources are available to ensure effective and reliable screening of at least 70% of the target age group, women over the age of 50 years. Factors that influence participation include the availability of local health care services, trust in health care providers, and the level of governmental or private health care coverage. This paper does not examine the human and political factors that may significantly affect acceptability, but does model operational factors such as the dynamics of capacity and waiting, as well as the relationship between the proximity of service facilities and the likelihood of participating in a screening program [25]. While not all regions experience problems with waits, some do, and limited service capacity or scheduling are operational issues that can cause waits of even 3-6 months [26, 27].

Delays in the screening and diagnosis system [28-31] can reduce survival rates by delaying the stage of disease at diagnosis [32]. Increasing the minimum annual screening volume for accreditation may aggravate the capacity problem by reducing the number of radiologists willing or eligible to provide the service, thereby reducing accessibility. On the other hand, increasing the number of readings would increase the accuracy for communities that are still served. The dynamic interactions of reading volume and quality, access, delays in service and health outcomes, among other complications, present a challenge for health care service system design.

This paper presents a system dynamics model of screening services that combines the interactions described above, and uses the model to analyze the impact of different interventions on the system performance in terms of health outcomes and costs. The general modeling framework is applicable to preventive health care services in general and is aimed at contributing to a better understanding of health care system design. Section 2 identifies papers that have studied some aspects of these important determinants of mammogram screening program success, but no paper seems to account for all of these potential interactions at once. It then presents a mathematical model of these interactions. The model is validated with data from published studies where possible, and simulation experiments in Section 3 are motivated by policy issues that arise in the U.S. and French health care contexts. A system dynamics model with deterministic differential equations might seem appropriate at first glance, but the model employs stochastic dynamics so that waiting can be more adequately described. Section 4 discusses implications and limitations of the model, as well as further research directions.

2   Problem Formulation

We view breast cancer screening provision as a problem of matching the supply (screening service) and demand (participation in the program, screening frequency) while ensuring sufficient quality (high sensitivity and specificity of the test). The objective is to reduce the breast cancer deaths, keeping system costs in mind.

There are several related health care service delivery papers in the operations management literature. Location models [7, 8] help situate screening facilities, but model neither the waiting list and health outcomes nor the documented relationship between screening volume and quality. [33] is unique in that it considers the deleterious health effects of waiting due to constrained health care service capacity. Their deterministic differential equation model is appropriate for the short-term transient dynamics of their application (smallpox control) but is less appropriate for assessing the long run performance of a continuously operating health care service. [34] models the interrelationship of service levels and system capacity, costs, and health outcomes

4

(for renal disease), and obtains structural results with a fluid model. Those papers do not include the stochastic effects and waiting that we model. [35] is related in that it models the deleterious health effects of delayed health service provision in the context of kidney transplant. That work applies more to sequential stochastic assignment problems, but does not model programs with repeated screening.

Several studies of breast cancer screening in the operations research literature focus on modeling the disease progression to optimize the screening schedules for an individual [10, 36-38]. Their goal is similar to one of ours: to evaluate performance of different screening policies to find one maximizing the health outcomes and/or minimizing the costs. Among different objectives of these models are minimizing costs associated with screening and the disease [37-39], minimizing detection delay given a number of screens [36] or more general utility functions related to detection time [10]. A limitation of these studies is that they do not consider the system-level constraints such as limited service capacity or delayed arrivals of women with a demand for screening. Moreover, they model the quality of screening test as a constant or as a function of tumour size, but do not incorporate the effect of radiologist experience. Our model addresses these points, while it takes the screening frequency as a given parameter.

Section 2.1 describes our choice for breast cancer disease progression and mammogram program service delivery structure models. Section 2.2 does the same for the relationship between screening volume and quality, and describes assumptions about acceptability. Section 2.3 gives cost assumptions.

2.1 Disease Progression and Service System Structure

We model two types of service (both screening tests and follow-up diagnostic tests), a finite service capacity, and the potential of waits due to finite service capacity and randomness associated with patient scheduling. There are n parallel servers (radiologists) that serve c queues (facilities).

Figure 2 depicts a system with $c = 1$ facility. A fraction h of individuals that reach the target screening program age (say, 50 year old women) join the screening program, the remainder are considered to be

unenrolled. Enrolling individuals have an early stage cancer with probability p. With frequency f, enrollees attempt to obtain a screening mammogram (a type $t = s$ job) but wait in queue if all servers are busy. The service time for screening mammogram is an exponential random variable with rate μ_S. The service rate represents the bottleneck resource in the facility. If the screening mammogram result is positive, a diagnostic test is required. This means a $t = s$ job will return to the queue as a type $t = d$ job with a probability that depends on the health state of the patient and sensitivity and specificity of the screening. Diagnostic test flows, represented with dashed lines, have a service rate $\mu_D = \mu_S/a$, a higher sensitivity than screening mammograms, and incur greater costs than screening mammograms. If cancer is detected after completion of diagnosis, the woman goes under treatment. Otherwise she goes back to the target population. To reflect exogenous sources of detection such as self-exam and general practitioner referral, diagnostic tests may also be requested directly without a screening mammogram (with rate m_1 and m_2 respectively for women with early and late stage cancer). Unenrolled individuals may enroll in the program later, with rate γ.

FIGURE 2 NEAR HERE

The numbers in Figure 2 identify the state of progress through the health service system, (1) unenrolled, (2) enrolled but not yet scheduled for a screening, (3) waiting for a screening mammogram, (4) getting a screening mammogram, (5) waiting for a diagnostic test, (6) getting a diagnostic test, or (7) undergoing treatment for cancer. We do not model treatment capacity explicitly, so a woman diagnosed with breast cancer begins treatment immediately.

We assume a three-stage health model like many others [5, 10, 36, 39]: (1) healthy, (2) preclinical, or early stage, breast cancer (3) clinical, or late stage, breast cancer. The number of individuals in health status $j = 1, 2, 3$ and service system state $i = 1, 2, \ldots, 7$ is denoted $X_{i,j}$. The service capacity constraint limits the number being served at any given time, $\sum_{j=1}^{3} X_{4,j} + X_{6,j} \leq n$.

Patient flows through the health system are represented vertically in Figure 3, and changes in health status are illustrated with horizontal flows from compartment to compartment. Table 1 gives the default values of

6

the parameters that determine the flows, including parameters introduced above as well those that describe screening quality introduced below. A description of how their values were validated relative to published data and statistical information from national agencies can be found in the Appendix. We assumed a fixed accuracy level for diagnostic follow-up tests, while the accuracy of screening mammograms is calculated as a part of the model, as described in the following section. The parameter values give good fit with incidence and breast cancer death data from Statistics Canada [40].

FIGURE 3 NEAR HERE

TABLE 1 NEAR HERE

We model two causes of death, disease specific mortality and all-cause mortality [41]. Noncancer related deaths occur at rate g and may affect women in all compartments (arcs for these transitions are not shown in the figures for clarity). Cancer-related deaths are assumed to affect only women with late stage cancer ($X_{i,3}$) or women undergoing treatment ($X_{7,j}$). A constant population size is assumed (a new individual enters the target population as a death occurs).

For a given sensitivity and specificity, the dynamics of disease progression, program enrollment, and screening outcomes are assumed to be Markovian. In addition to the $X_{i,j}$, the state includes information about the volume of screening mammogram readings of the servers, which influences the reading quality, as described in Section 2.2. Quality influences the system dynamics model in Figure 3 via the flows on the thick arrows that are associated with false positive ($X_{4,1}$ to $X_{5,1}$ and $X_{6,1}$ to $X_{7,1}$) and false negative ($X_{4,2}$ and $X_{6,2}$ to $X_{2,2}$, and $X_{6,3}$ to $X_{2,3}$) test results.

2.2 Volume-Quality Relationship and Acceptability

Many factors influence the quality of readings [17, 24, 43], and acceptability (measured here by the fraction of women that enroll in a breast cancer screening program). Here we focus on the relationship between volume

and quality. The quality-volume relationship is uncertain and complex, and its accuracy is questioned by some articles [44, 45]. We rely on the literature that supports this relationship [23, 46] to provide a model for quality of mammogram reading. We then describe a simple model for acceptability.

Sensitivity and specificity increase as the average reading volume increases [23, 46]. We assume a logistic relationship between volume and quality that embodies qualitative features from empirical observations [46].

$$\text{sensitivity} \quad = \quad \alpha(v) = \frac{0.95}{1 + 0.5393e^{-0.0021v}}$$

$$\text{specificity} \quad = \quad \beta(v) = \frac{0.95}{1 + 0.158e^{-0.0024v}}$$

where v is a monthly reading volume. These functions are plotted in Figure 4.

FIGURE 4 NEAR HERE

There are several choices for modeling the volume of readings per server. We chose to measure the screening volume for fixed-length time periods and base quality during one time on the volume in the preceding time period. Specifically, if $vol(t)$ is the total number of completed mammogram readings up to time t years, then the volume of readings during the previous month is

$$v(t) = vol\left(\frac{\lfloor 12t \rfloor}{12}\right) - vol\left(\frac{\lfloor 12t \rfloor}{12} - 1\right).$$

An alternative approach to model quality-volume relationship could be a discrete state Markovian model (with knowledge measured by the number of readings recalled and that are effective for quality output, $v(t) \in \{0, 1, 2, \ldots\}$, with transition $v \to v + 1$ upon service completion, and $v \to v - 1$ with forgetting rate αv, where $1/\alpha$ is the 'half life' of recalling a screen). This approach leaves a memory of readings that is highly volatile. A similar continuous state, continuous time model for $v(t)$ might also be easier to study analytically, but also neglects the issue of regulatory checks which count readings in specific fixed length time periods. Formally, our choice to measure volume in fixed-length time periods changes the stochastic process in Section 2.1 from a Markov chain to a generalized semi-Markov process. We use simulation for the analysis, a standard tool for studying this class of processes.

Acceptability, measured by the probability h that a woman initially enrolls in a screening program, is simplified here to depend upon the distance from the nearest facility. We set h to match national enrollment statistics in some experiments. In others that test the effect of distributed facilities versus a centralized facility (requiring longer travel), we assume the odds ratio of enrolling drops by 3% for each additional 8km traveled, as in an empirical study [25]. This distance-enrollment relation oversimplifies a complex set of effects, but like the quality-volume relationship, it is probably the best we can use on the basis of research now available. Other factors that involve screening quality and acceptability can be modeled similarly.

2.3 Cost Assessment

The economic costs of screening, diagnostic follow-up tests and treatment in Table 2 are based on values taken from the literature and converted to 2003 U.S. dollars using the consumer price index for medical services where necessary. The cost of follow-up diagnosis tests is based on a weighted average of the costs of diagnostic mammogram, sonography, fine needle aspiration and biopsy reported in [47]. Long term discounted costs of treatment and continuing care are based on a three stage model for breast cancer classification [48]: local, regional or distant cancer. Local cancer corresponds to our preclinical stage. Both regional and distant stages correspond to our clinical stage. We therefore use a weighted average late treatment cost for the regional and distant stages, $\$68,551 = (\$70,066 \times 0.3 + \$59,463 \times 0.05)/0.35$. This lets us use incidence data to calculate the expected costs of breast cancer cases. While the cost conclusions below are based on data available from one HMO, some generality is preserved because the relative costs of different tests and treatment at different stages are more relevant for our purposes than the absolute figures, and cost comparisons with other previous studies do not indicate a sizable difference [48].

<div align="center">**TABLE 2 NEAR HERE**</div>

Like [47, 14, 16], we do not explicitly account for the cost of increasing program enrollment, service capacity, or changing screening standards. Those costs are likely to be highly dependent upon the target

population. [49, 50] illustrate how to include such costs. We also vary the program enrollment probability over a wide range for sensitivity analysis. Realistically it may be expensive or even impossible to achieve very high, or even very low, enrollment levels.

3 Analysis

This section presents four analyses motivated by issues in the U.S. and French health systems in three subsections. The first experiment assesses the cost implications for two approaches to improving early detection, either through outreach or through quality increases due to increasing the minimum screening volume standards. The second experiment examines interactions between quality and the potentially deleterious health effects of waiting in the presence of insufficient capacity. The last two examine the interactions of service decentralization, access, and screening quality. Since the model is not easy to analyze in closed form, simulation experiments were used to estimate long run averages with batch means [51] for health outcomes and annual costs. Tests for stationarity with the Heidelberger and Welch [52] test led us to remove 20 years of 'warm-up' from the beginning of each simulation of 400-520 years for each parameter setting.

3.1 Increasing standards or expanding outreach

The National Cancer Institute (NCI) recommends mammography screening every one to two years for American women over 40 [53]. The General Accounting Office [54] estimates that 2/3 of the mammography machine capacity is utilized and that 64% of the target population had a screening mammogram in 2000, less than the 70% recommended by the WHO. Waiting is not significant on the whole, although waits of several months occur in some metropolitan and rural areas. The U.S. FDA currently requires radiologists to interpret a minimum of 480 screenings per year [22]. If participation increases to 70%, more cancers will be detected early both because more women are being screened and because screening quality improves with an increased volume per radiologist. On the other hand, an increase in demand along with recent decreasing

10

trends in capacity [54] may exacerbate waiting and lessen the ability to detect cancers early. An alternative to increasing participation directly is to improve screening quality, and therefore health outcomes, by increasing the minimum annual screening standard from 480 to 2, 500 (the figure recommended by [23]).

This section examines the following questions. What health benefits can be gained by increasing outreach and what would be the impact for the capacity requirements? What are the benefits of increasing the minimum screening volume to 2, 500 per year? What are the implications on capacity requirements?

We simulated a target population of 25, 000 women. Since readings are typically not the only service provided [44], we simulated both the 480 base-level screening standard and increased reading standard of 2, 500/year by presuming that the maximum rate of readings would be about twice the standard level (so $\mu_S = 1, 000$ for base-level screening, and $\mu_S = 5, 000$ for the increased level). Initially, the fraction enrolling immediately is $h = 0.55$, so that the long run participation in the screening program roughly matches the empirical 64% value [54] (some women enroll later spontaneously or upon noticing symptoms). To evaluate the effect of increasing outreach and acceptability, we checked multiple h from 0.55 to 0.75 for both scenarios. The number of radiologists is set so that 30, 000 mammogram screenings per year can be done. Table 3 summarizes the parameters used in the experiments.

TABLE 3 NEAR HERE

Health Outcomes. Figure 5 shows that the average number of early diagnoses per year increased when the reading standards were increased to 2, 500 from 480, regardless of the level of participation in the screening program. This was a direct result of the improved sensitivity and specificity of readings at higher volumes. Figure 5 also shows that for a given volume standard, increased participation levels resulted in increased early detection. This was a compound effect due to more women being screened and higher quality of readings due to a higher volume per radiologist.

FIGURE 5 NEAR HERE

11

An increase in the number of early diagnoses was reflected in a decrease in the breast cancer mortality results. Increasing the reading volume standard to 2, 500 at 65% participation level had approximately the same effect on the breast cancer death rate as increasing the participation to 69% (a 2-3 % decrease in the number of breast cancer deaths). In order to understand the relative costs and benefits of increasing outreach versus increasing quality, we compared these two specific options. Table 4 summarizes the parameters for the numerical experiments, and Table 5 reports the results (intervals represent 90% confidence intervals). With both options, an equivalent improvement in health outcomes was achieved.

TABLE 4 NEAR HERE

TABLE 5 NEAR HERE

Cost of Screening and Treatment. Table 6 summarizes the costs of each program, combining the screening and diagnostic tests and treatment costs from Section 2.3 and the outcomes in Table 5, assuming the cost of treating after a false positive is the same as the cost of treatment after early diagnosis. For example, the estimated total annual cost of screening and treatment for option 1 included the costs of screening mammograms, diagnostic follow-up tests, and treatment for early and late stage cancers, and false positive diagnosis: $17,132 \times \$145 + 2,909 \times \$471 + 27.4 \times \$54,013 + 53.3 \times \$68,551 + 133 \times \$54,013 = \$16.0 \times 10^6$.

TABLE 6 NEAR HERE

Increasing the reading volume standards (Option 2) resulted in costs $2,600,000 less than option 1, while providing equivalent health outcome benefits because of two effects. An improvement in specificity decreases the unnecessary diagnostic procedures. An improvement in sensitivity increased the chances of detecting an actual tumor. These quality improvements are desirable for both costs and health outcomes. On the other hand, increasing outreach while keeping the standards the same (Option 1) increased costs by increasing the total screening costs and the number of unnecessary diagnostic mammograms in order to achieve comparable

12

health benefits (Figure 6 compares the number of false positive test results). The combined impact of these options 1 and 2 would be a further reduction in breast cancer deaths (4.7%), with an estimated total cost of 14.3×10^6. The marginal cost increase due to increased outreach is therefore lower when quality standards are higher ($16.6-16.0 = $0.6 > $14.3-$14.0=$0.3) because there are fewer unneeded diagnostic tests.

FIGURE 6 NEAR HERE

The costs of achieving these improvements are *not* included in calculations, since they depend on the specific health care context and require capacity investment. Increasing participation to $69 - 70\%$ may be expensive. Small improvements on both participation and quality will be preferable to increasing one or the other if improvement costs are convex.

This result is not a call for not increasing the outreach of screening programs, but a warning for the costs of low quality screening. Increasing outreach provided substantial health outcome benefits and is desirable in order to provide an egalitarian public health service. If the quality of the screening test were low, by expanding outreach there would be excess waste in the system and the costs would increase unproportionally with the health outcome benefits. Further, the benefits from screening more women were not fully realized when the standard (that is, the quality) was low, since a high percentage of the early stage cancers were missed among the ones screened.

Capacity Requirements. The simulation results did not indicate a problem with insufficient capacity and waiting up to a participation rate of 81%. Waiting times were not significant and did not affect health outcomes. Comparing the capacity requirements of the two options reinforced the benefits of increasing reading volume standards (Option 2). The higher quality due to the higher standards level reduced the load on the system that resulted from diagnostics required to resolve false positive results (Figure 6). Consequently, when the quality was low, the utilization level was always higher due to that indirect effect on the system load, so increased capacity requirements are a more serious problem with lower quality readings. The degree to which increased waits may negatively affect health outcomes is explored in Section 3.2. The greater

the resource needed for diagnostic tests (larger *a*), the greater the capacity constraint effect caused by false positives. Decoupling screening from diagnostic mammogram capacity would reduce that effect.

The above comparisons are based on the costs of screening, diagnosis and treatment. They do not include patient-related costs like anxieties associated with false positives, or the effects of false positive results and long waiting lists on the willingness of women to request screening. If those additional factors were considered, the advantage of the option of improving quality over the option of increasing outreach would be even more significant.

These observations require some caveats. We focused on the effect of increasing the standards for reading volume on quality in our experiments. There can be some other consequences of increasing the standards. Fewer doctors may be willing to dedicate a significant proportion of their time to mammogram reading with higher demand levels. As the number of eligible doctors decreases, participation may decrease too since the transportation times will increase. Section 3.3 explicitly accounts for the participation and distance effect in a separate experiment. Finally, increasing outreach improves the chance of early detection to a broader cross-section of women, and may influence program design decisions on ethical grounds.

3.2 Limited Capacity, Waits, and Delayed Detection

Capacity crises may occur if demand increases and/or capacity decreases. While this may not be the case globally, waits occur in some areas [54], and many countries plan to increase participation. In the UK, women aged between 50-64 are screened, but work is being carried out to extend invitations to women up to age 70 by 2004 [55]. France intends to improve breast cancer screening participation to 80% of the target population by 2007 [56]. While these extension plans are implemented, capacity implications should be considered carefully, since capacity may be slower to influence because of extensive training required.

We ran simulations with the input parameter values in Table 7 to explore the relationship between capacity, utilization, waiting, and health outcomes. The recommended screening interval differs from country to

country. Here we set it to 2 years.

TABLE 7 NEAR HERE

Figure 7 shows how insufficient capacity counteracts the benefits expected from increasing participation. This happened when the additional demand was not met and long waiting lines were observed. Figure 7 shows that there was a decrease in the number of cancer deaths as participation increased to 65% (corresponding to a utilization of 99%). Additional demand increased congestion and women had to wait to get regular screening mammograms. The output is given in Table 8. In this experiment, when participation is 96%, the average waiting time is 8.5 months (average waits by Little's Law are $6873/9670 = 0.71$ year). Capacity constraints or other causes for several months delay beyond a two-year screening interval can lead to poorer health outcomes due to fewer early detections. These deleterious health effects can be mitigated by improving quality for two reasons. First, each test is more accurate, improving detection. Second, a reduced burden due to less frequent follow-up diagnostic tests can free up capacity to further reduce waiting.

FIGURE 7 NEAR HERE

TABLE 8 NEAR HERE

Additional runs with no service burden due to diagnostic tests ($a = 0.001$) indicated qualitatively the same result, with a small twist. A decrease in mammogram resource requirements for diagnostic tests increases the optimal participation rate. The minimum breast cancer death rate was obtained at 81% utilization.

Our results suggest that waiting would affect health outcomes only when there is a severe capacity problem. Although screening mammograms are planned and scheduled, there is a tendency to stretch out the screening intervals, a phenomenon called "slippage" ([57] reported in [27]). Waiting cannot be avoided completely even when there is organized screening and scheduling in place. It is therefore important to consider the stochastic aspects of the demand for screening and the fact that schedules may not be implemented

as intended. When capacity is insufficient, the problem will be aggravated and will have adverse effects on health outcomes.

3.3 Decentralization Decision / Learning From Peers

This section models one factor that influences participation: the use of decentralized facilities to reduce the distance traveled to the nearest facility [25], operationalized by mobile clinics or putting equipment in the facilities of more primary care providers, and modeled by the distance/access relationship in Section 2.2. Decentralization may have a positive effect in that improved participation offers more chances of early detection, and increases volume and quality. On the other hand, more facilities implies lower volume per facility. If quality is improved in centralized facilities because outlier results can be shared with peers, this effectively improves the volume that each colleague sees. Reading quality in a centralized facility will be somewhere in between the quality that corresponds to the volume seen working alone, and the total volume of a centralized facility. As a result, decentralization may have mixed effects on reading quality while increasing the participation rates. The net effect is unknown [58].

We consider four cases with respect to two factors: (1) the effect of decentralization on quality (with learning from peers in a centralized facility, or without learning) and (2) capacity (sufficient capacity exists or not). If there is learning with centralization, we assume the best possible case, that the quality of each individual radiologist is based on the total volume of readings at the facility. Without learning, quality is modeled as before, as a function of the individual reading volume.

We assume that 60,000 women are evenly distributed over 100km, and go to the nearest of $c = 1, 2, 4$ or 8 facilities, assumed to be evenly distributed. The facilities house a total of 8 radiologists, each of whom serves at a rate of $\mu_S = 5000$/year. To model the sufficient and insufficient capacity cases, we set the maximum enrollment probability (with no travel) to $h_0 = 0.35$ and $h_0 = 0.75$ respectively. Participation rates ranged from $44 - 49\%$ for $h_0 = 0.35$ (sufficient capacity) and from $77 - 80\%$ for $h_0 = 0.75$ (insufficient capacity).

With learning from peers, the volume associated with a fully centralized facility ($c = 1$) corresponds here to a sensitivity and specificity of about 0.95, which represents an upper bound quality level. When there is no learning, average sensitivity ranges in 0.78-0.83 while specificity is about 0.89.

Learning With Centralization. When quality is attenuated because centralization is associated with better reading performance, then Figure 8 indicates that the value of decentralization depends upon whether there is sufficient capacity to meet demand or not. If the system is already under-capacitated, then the fraction of the population actually screened may decrease, in spite of the fact that more people seek screening. Decreased reading quality in a decentralized setting increased demand for follow up tests due to false positives, reducing the effective capacity for screening mammograms. On the other hand, if there was sufficient capacity, then decentralization increased the ability to screen more women. The right panel of Figure 8 indicates that the net effect on annual breast cancer deaths was more complicated. Initially, decentralization reduced cancer deaths, due to early detection for more women. The benefits of increasing participation outweighed the losses in quality and pooling efficiency. But too much decentralization decreased reading quality and missed early stage cancers. Moreover, a loss of pooling advantage further increased waiting times and decreased the chance of early detection, so breast cancer deaths started to increase again. For the fully centralized (1 facility) and fully decentralized (8 facilities) cases, the number of breast cancer deaths were at about the same level. This suggested that learning in a centralized facility (which increased the sensitivity from 0.77 to 0.94) could provide the same benefits as decentralization, which increased participation from 44% to 49%. When there was insufficient capacity, the results do not suggest that decentralization decreased breast cancer deaths in the same way.

FIGURE 8 NEAR HERE

No Learning with Centralization. If quality is not affected by decentralization, we observed less impact of decentralization on the percent population screened, because the 'false positives effect' was weaker. By increasing the number of facilities, the percent of the population screened remained constant when there was

insufficient capacity. The fraction screened increased when there was sufficient capacity (left panel of Figure 9). The effect on annual breast cancer deaths followed a similar pattern: Since there is no loss in quality, when there is sufficient capacity the death rate decreased. Decentralization had no statistically significant effect on death rates when there was insufficient capacity because resources were already fully utilized (right panel of Figure 9).

FIGURE 9 NEAR HERE

The model suggests that fixed costs aside, decentralization is advantageous up to the point where screening quality drops significantly. If quality can be maintained in decentralized facilities, decentralization is beneficial as long as there is enough capacity to meet the increased demand. If decentralization is not an option for other reasons, then instituting practices that enhance learning in centralized facilities can provide most of the reduction in cancer deaths that decentralization can provide.

4 Discussion

Our stochastic system dynamics model includes several factors that have not yet been considered all at once in the mammogram screening literature. Simulations here illustrated the system behavior, health outcomes and costs for some aspects of breast cancer screening programs due to public policy actions like improving enrollment rates or quality standards for radiologist certification.

A similar approach can be useful in other applications like colorectal cancer screening, where volume and quality; demand and the degree of facility decentralization; or capacity, service delays and outcome quality are interrelated. Colonoscopy is widely viewed as the most accurate screening test for colon cancer, and demand for colonoscopy has surged so much in recent years that patients may wait for months or be turned away [59]. Service design issues for colonoscopy also include the use of multiple screening policies with different costs, sensitivities and specificities.

18

The experiments here highlight the importance of the sensitivity and specificity dimensions of screening quality. Low quality results in additional follow-up tests that waste system capacity. The U.S., France and other countries have plans to increase adherence to regular screening by decentralization or other means, and many regions are experiencing a decline in service capacity. Any increase in participation should be accompanied both by an assurance that sufficient capacity will be established, and a maintenance or increase in screening quality to insure that delays due to system dynamics do not decrease or reverse the anticipated public health benefit. Low reading volume standards reduce the quality of readings and increase screening costs by increasing the workload due to follow-up tests. Health outcomes could deteriorate because of a decreased effectiveness of screening and potential delays that might result in a late diagnosis. Decentralization of screening service to increase participation in screening is found beneficial only if the quality of screening tests can be maintained. These interactions between volume, quality, capacity and waiting influence health outcomes and system costs in ways that have not fully been accounted for in previous studies.

These aggregate conclusions should be understood relative to the limitations of the model. The homogeneous population assumption ignores risk factors involving age, genetic disposition, and environmental effects. Scheduling can ideally reduce waiting times but cannot prevent waiting completely because of the compliance issues discussed in Section 3.2, so we did not consider it here. Since health effects are primarily deleteriously affected by waiting times when capacity is insufficient, the value of scheduling would appear to be a second-order effect. The three-stage health model does not focus on tumor growth dynamics and patient-to-patient variability, but is consistent with a number of other papers. Quality was assumed to be a function of screening volume alone here. Adjustments can be made to handle other features [24, 60], like age and variability in reading quality between doctors, other skill factors, film quality, and controllable trade-offs between specificity and sensitivity in reading assessments, but we did not do so here.

Screening and treatment costs are included, but not the cost of the improvement options. Those must be added based on the specific health care environment to obtain a full cost-benefit analysis and to better inform

controversy over the real value of breast cancer screening. Our aggregate level model did not focus on the incentives of each actor in the health care system. The incentives of patients, providers and payers also play a role in determining service capacity and participation rate figures. Poor insurance coverage decreases the willingness of women to participate. Low reimbursement rates and high certification standards may decrease the willingness of the radiologists to provide service, in favor of other more profitable tasks. These issues could be explored with suitable data.

References

[1] Imaginis. Breast cancer: Statistics on incidence, survival and screening. 2002. `imaginis.com/breasthealth/statistics.asp#1`, retrieved on 17/05/2004.

[2] ACS: Cancer facts and figures. (American Cancer Society, Atlanta, GA 2003).

[3] H. Thornton, A. Edwards, and M Baum. Women need better information about routine mammography. British Medical Journal 327 (2003) 101–103.

[4] C. Klabunde, F. Bouchard, S. Taplin, A. Scharpantgen, and R. Ballard-Barbash. Quality assurance for screening mammography: an international comparison. Journal of Epidemiology and Community Health 55 (2001) 204–212.

[5] H. Lee and W. Pierskalla. Mass screening models for contagious diseases with no latent period. Operations Research 36 (1998) 917–928.

[6] S. Ozekici and S. Pliska. Optimal scheduling of inspections: A delayed Markov model with false positives and negatives. Operations Research 39 (1991) 261–273.

[7] S. Lapierre, D. Ratliff, and D. Goldsman. The delivery of preventive health services: A general model. (Technical report, Georgia Institute of Technology, 1997).

[8] V. Verter and S. Lapierre. Location of preventive health care facilities. Annals of Operations Research 110 (2002) 121–130.

[9] N.E. Day and S.D. Walter. Simplified models of screening for chronic disease: Estimation procedures from mass screening programs. Biometrics 40 (1984) 1–14.

[10] M. Zelen. Optimal scheduling of examinations for the early detection of disease. Biometrika 80 (1993) 279–293.

[11] S.W. Duffy, H.H. Che, L. Tabar, and N.E. Day. Estimation of mean sojourn time in breast cancer screening using a Markov chain model of both entry to and exit from the preclinical detectable phase. Statistics in Medicine 14 (1995) 1531–1543.

[12] J. Xu, M.R. Fagerstrom, and P. Prorok. Estimation of post-lead-time survival under dependence between lead-time and post-lead-time survival. Statistics in Medicine 18 (1999) 155–162.

[13] P. Nutting, D. Iverson N. Calonge, and L. Green. The danger of applying uniform clinical policies across populations: The case of breast cancer in American Indians. American Journal of Public Health 84 (1994) 1634–1636.

[14] F. Boer, H. de Koning, P. Warmerdam, A. Street, E. Friedman, and C. Woodman. Cost effectiveness of shortening screening interval or extending age range of NHS breast screening programme: Computer simulation study. British Medical Journal 317 (1998) 376–379.

[15] P.M. Clarke. Cost benefit analysis and mammographic screening: A travel cost approach. Journal of Health Economics 17 (1998) 7367–787.

[16] D. Gyrd-Hansen and J. Søgaard. Analyzing public preferences for cancer screening programs. Health Economics 10 (2001) 617–634.

[17] W.A. Berg, C.J D'Orsi, V.P. Jackson, L.W. Bassett, C.A. Beam, R.S. Lewis, and P. Crewson. Does training in the breast imaging reporting and data system (BI-RADS) improve biopsy recommendations or feature analysis agreement with experienced breast imagers or mammography? Radiology 224 (2002) 871–880.

[18] J.P. Caulkins and G. Tragler. Dynamic drug policy: an introduction and overview. Socio-Economic Planning Sciences 38 (2004) 1–6.

[19] S.E. Chick, S. Soorapanth, and J.S. Koopman. Microbial risk assessment for drinking water. in Operations Research and Health Care: A Handbook of Methods and Applications, ed. M. Brandeau, F. Sainfort, and W. Pierskalla, (Kluwer Academic Publishers, 2004).

[20] World Health Organization. Screening for various cancers. 2003. `www.who.int/cancer/detection/breastcancer/en/`, retrieved on 17/05/2004.

[21] M. Moss. Spotting breast cancer: Doctors are weak link, NY Times (2002).

[22] Mammography Quality Standards Act regulations. FDA Website. 2002. `www.fda.gov/cdrh/mammography/frmamcom2.html#s90012`, retrieved on 8/5/2004.

[23] A. Kan, I. Olivotto, L.W. Burhenne, E. Sickles, and A. Coldman. Standardized abnormal interpretation and cancer detection ratios to assess reading volume and reader performance in a breast screening programme. Radiology 215 (2000) 563–567.

[24] J.G. Elmore, D.L. Miglioretti, L.M. Reisch, M.B. Barton, W. Kreuter, C.L. Christiansen, and Fletcher S. Screening mammograms by community radiologists: variability in false-positive rates. Journal of the National Cancer Institute 94 (2002) 1373–1380.

[25] K. Engelman, D.B. Hawley, R. Gazaway, M.C. Mosier, J.S. Ahluwalia, and E.F. Ellerbeck. Impact of geographic barriers on the utilization of mammograms by older rural women. American Geriatrics Society 50 (2002) 62–68.

[26] J. Fischman. New-style mammograms detect cancer. So do the old. Either way you wait. 2001. `nl.newsbank.com/`, retrieved on 08/05/2004.

[27] Organized breast cancer screening programs in Canada: 1997 and 1998 report. (Health Canada, 1998).

[28] P. Thongsuksai, V. Chongsuvivatwong, and H. Sriplung. Delay in breast cancer care: a study in Thai women. Medical Care 38 (2000) 108–114.

[29] M. Montella, A. Crispo, G. Botti, M. De Marco, G. Bellis, G. Fabbrocini, M. Pizzorusso, M. Tamburini, and G. Daituo. An assessment of delays in obtaining definitive breast cancer treatment in southern Italy. Breast Cancer Research and Treatment 66 (2001) 209–215.

[30] L.S. Caplan, K.J. Helzlsouer, S. Shapiro, L.S. Freedman, R.J. Coates, and B.K. Edwards. System delay in breast cancer in whites and blacks. American Journal of Epidemiology 142 (1995) 804–812.

[31] L.S. Caplan, K.J. Helzlsouer, S. Shapiro, M.N. Wesley, and B.K. Edwards. Reasons for delay in breast cancer diagnosis. Preventive Medicine 25 (1996) 218–224.

[32] M.A. Richards, A.M. Westcombe, S.B. Love, P. Littlejohns, and A.J. Ramirez. Influence of delay on survival in patients with breast cancer: A systematic review. The Lancet 353 (1999) 1119–1126.

[33] E.H. Kaplan, D.L. Craft, and L.M. Wein. Emergency response to a smallpox attack: The case for mass vaccination. PNAS: Proceedings of the National Academy of Sciences 99 (2002) 10935–10940.

[34] S.A. Zenios and P.C. Fuloria. Managing the delivery of dialysis therapy: A multiclass fluid model. Management Science 46 (2000) 1317–1336.

[35] X. Su and S.A. Zenios. Allocation of kidneys to autonomous transplant candidates: A sequential stochastic assignment model 2004. submitted to Operations Research.

[36] R.L.A. Kirch and M. Klein. Surveillance schedules for medical examinations. Management Science 20 (1974) 1403–1409.

[37] M. Schwartz. A mathematical model used to analyze breast cancer screening strategies. Operations Research 26 (1978) 937–955.

[38] R. Baker. Use of a mathematical model to evaluate breast cancer screening policy. Health Care Management Science 1 (1998) 103–113.

[39] J. Voelker and W. Pierskalla. Test selection for a mass screening program. Naval Research Logistics Quarterly 27 (1980) 43–56.

[40] Cancer surveillance on-line. 2003. `dsol-smed.hc-sc.gc.ca/dsol-smed/cancer/index_e.html`, retrieved on 17/05/2004.

[41] W.C. Black, D.A. Haggstrom, and H.G Welch. All-cause mortality in randomized trials of cancer screening. Journal of the National Cancer Institute 94 (2002) 167–173.

[42] S.D. Walter and N.E. Day. Estimating the duration of a pre-clinical disease state using screening data. American Journal of Epidemiology 118 (1983) 865–886.

[43] C.F. Nodine, H.L. Kundel, C. Mello-Thoms, S.P. Weinstein, S.G. Orel, D.C. Sullivan, and E.F. Conant. How experience and training influence mammography expertise. Academic Radiology 6 (1999) 575–585.

[44] C. Beam, E. Conant, and E. Sickles. Factors affecting radiologist inconsistency in screening mammography. Academic Radiology 9 (2002) 531–540.

[45] J.G. Elmore, D.L. Miglioretti, and A.P. Carney. Does practice make perfect when interpreting mammography?, part ii. Journal of the National Cancer Institute 95 (2003) 250–252.

[46] L. Esserman, H. Cowley, C. Eberle, A. Kirkpatrick, S. Chang, K.Berbaum, and A. Gale. Improving the accuracy of mammography: Volume outcome relationship. Journal of the National Cancer Institute 94 (2002) 369–375.

[47] P. Salzman, K. Kerlikowske, and K. Phillips. Cost-effectiveness of extending screening mammography guidelines to include women 40 to 49 years of age. Annals of Internal Medicine 127 (1997) 955–965.

[48] B.H. Fireman, C. Quesenberry, C. Somkin, A. Jacobson, D. Baer, D. West, A. Potosky, and M. Brown. Cost of care for cancer in a health maintenance organization. Health Care Financing Review 18 (1997) 51–76.

[49] M.R. Andersen, M. Hager, C. Su, and N. Urban. Analysis of the cost-effectiveness of mammography promotion by volunteers in rural communities. Health Education and Behavior: The Official Publication of the Society for Public Health Education 29 (2002) 755–770.

[50] R. Saywell, V. Champion, T. Zollinger, M. Maraj, C. Skinner, K. Zoppi, and C. Muegge. The cost effectiveness of 5 interventions to increase mammography adherence in a managed care population. The American Journal of Managed Care 9 (2003) 33–44.

[51] A.M. Law and W.D. Kelton. Simulation Modeling and Analysis. (McGraw-Hill, New York, 2000).

[52] P. Heidelberger and P. Welch. Simulation run length control in the presence of an initial transient. Operations Research 31 (1983) 1109–1144.

[53] National Cancer Institute. NCI statement on mammography screening: 1/31/2002 update. `www.nci.nih.gov/newscenter/mammstatement31jan02`. retrieved on 6/7/2004.

[54] U.S. General Accounting Office. Mammography capacity generally exists to deliver services (GAO-02-532, Washington, DC, 2002).

[55] National Health Service (UK). Cancer screening programmes. 2002. `www.cancerscreening.nhs.uk/breastscreen/index.html#how-org`, retrieved on 17/05/2004.

[56] C. Haigneré J-F. Mattei. Cancer: Une mobilisation nationale, tous ensemble. 2003. `www.sante.gouv.fr/htm/dossiers/cancer/index2.htm`, retrieved on 17/05/2004.

[57] A.M. Faux, G.M. Lawrence, M.G. Wheaton, C.L. Jeffery C.L., and R.K. Griffiths. Slippage in the NHS breast screening programme: An assessment of whether a three year screening round is being achieved. Journal of Medical Screening 5 (1998) 88–91.

[58] U.S. Health Program Office of Technology Assessment. Screening mammography in primary care settings: Implications for cost, access and quality. J. Wagner, study director (1991).

[59] G. Kolata. 50 and ready for a colonoscopy? Doctors say wait is often long. NY Times (2003).

[60] E.A. Sickles, D.E. Wolverton, and K.E. Dee. Performance parameters for screening and diagnostic mammography: Specialist and general radiologists. Radiology 224 (2002) 861–869.

[61] Canadian cancer statistics. Toronto, Canada 2003. National Cancer Institute of Canada.

[62] W.E. Barlow, C.D. Lehman, Y. Zheng, R. Balard-Barbash, B.C. Yankaskas, G.R. Cutter, P.A. Carney, B.M. Geller, R. Rosenberg, K. Kerlikowske, D.L. Weaver, and S. Taplin. Performance of diagnostic mammography for women with signs or symptoms of breast cancer. Journal of the National Cancer Institute 94 (2002) 1151–1159.

[63] R. Pijnappel, M. van den Donk, R. Holland, W.P. Mali, J.L. Peterse, J.H.C.L. Hendriks, and P.H.M. Peeters. Diagnostic accuracy of different strategies of image guided breast intervention in cases of nonpalpable breast lesions. British Journal of Cancer 90 (2004) 595–600.

[64] H.M. Verkooijen, P.H. Peeters, E. Buskens, V.C. Koot, B. Rinkes, and T.J. van Vroonhiven. Diagnostic accuracy of large-core needle biopsy for nonpalpa breast disease: A meta analysis. British Journal of Cancer 82 (2000) 1017–1022.

A   Appendix: Parameter Estimates, Model Validation and Transition Rates

Table 1 summarizes the default values for parameters. They were estimated from medical journal articles or national statistical publications wherever possible to improve model validity. Where that was not possible, we made reasonable assumptions (b_3, b_{72}, b_{73}) or fit parameters (g, r_1, p_1, r_2, p_2, γ, m_1, δ, a) so that the simulation output was of the same magnitude as corresponding country statistics taken from Canadian Cancer Surveillance On-Line [40] (Table 9). [32] reports delay data for the time to apply for diagnosis after developing symptoms. We used the aggregate data to obtain an average delay of 2.9 months, or $m_2 = 4.13$/year. The probability of developing cancer per year is 2.5% for ages $50 - 59$, 3.1% for ages $60 - 69$, and 3.3% for ages $70 - 79$ (NCI of Canada [61] for 1998). We averaged the instantaneous rates of developing cancer for these age ranges to get $s_{12} := 0.0030122$. The 5 year survival probabilities for different cancer stages are taken from ACS data [2] in Table 10, so $b_{72} := 0.0081$ and $b_3 = b_{73} = 0.0915$ are weighted averages from the regional and distant categories.

A wide range of estimates for p_2 might be justified, and precise estimates of r_1, p_1, r_2 are not yet available. We therefore set the default values of these parameters to match flows. Death rate estimations are done using 5-year mortality figures, so we set $r_1 = 1/5$ and $p_1 = 0.05$ to get a recurrence rate of $r_1 p_1 = 0.01$ from early stage treatment. We assumed the recurrence rate tripled after treatment for late stage cancer, with $r_1 = 1/5$, $p_2 = 0.15$, so that $r_2 p_2 = 0.03$. That same product $r_2 p_2$ is obtained if $r_2 = 0.05$, $p_2 = 0.6$. Runs with the latter values would place a slightly higher screening load on the mammogram facility and a slight increase in waiting times, but wait times were not a significant deleterious factor in Section 3.1, so the results would differ litte with those parameter values. The rate r_0 of treatment completion after a false

positive diagnostic was set very high to model the continuing potential for the onset of preclinical cancer.

We assumed that the probability of joining the program later, and of asking for a diagnosis out of the screening schedule at an early stage of cancer, are small, so we set $\gamma = m_1 = 0.01$. Diagnostic follow up tests may include one or more of the diagnostic mammogram, fine needle aspirations, sonography, or biopsy [47]. Sensitivity estimates are 0.858 for diagnostic mammogram [62], 0.95 for fine needle aspiration [63], and 0.97 for biopsy [64]. As an average, sensitivity and diagnostic follow-up tests were set to 0.90 for preclinical cancer. In some rare cases, diagnostic tests may miss cancers, so we set the sensitivity of diagnostic test to 0.99 for clinical cancer. Specificity of diagnostic test is set to 0.95. The probability of having cancer at the entry to the target population is estimated using the data from Canadian Organized Breast Cancer Screening Program: cancer detection rate at first screen is 4.4/1000. With a sensitivity of 0.80 we get $p = 0.0055$.

TABLE 9 NEAR HERE

TABLE 10 NEAR HERE

Discrete changes for health status and position in the service system are essentially Markovian in continuous time, conditional on the reading quality, which may also vary through time. The quality-volume relationship is modeled as in Section 2.2. Quality, as measured by sensitivity and specificity, influence the type of transition when a screening mammogram is performed. Overall, there are 10 reasons for state changes and each occurs with the instantaneous transition rates given in Table 11, where X_{ij} represents the size of the compartment (i, j). The rates are sums, each summand representing flows out of individual compartments.

TABLE 11 NEAR HERE

Table 1: Summary of Default Values for Parameter Estimates

parameter	description
$g = 0.04$	death rate/person/year for reasons other than breast cancer (implied conditional life expectancy is 75 years: 50 years upon entry to target population, plus $1/g = 25$ years)
$\gamma = 0.01$	screening program enrollment rate/person/year for unenrolled
$p = 0.0055$	probability of having cancer at the entry to the target population
$b_3 = 0.0915$	probability of death from late stage cancer when not having treatment/person/ year
$b_{73} = 0.0915$	death rate/person/year from late stage cancer during treatment
$b_{72} = 0.0081$	death rate/person/year from early stage cancer during treatment
$s_{12} = 3.0122 \times 10^{-3}$	rate/person/year of acquiring preclinical cancer
$s_{23} = 0.585$	rate/person/year of cancer advancing from preclinical to clinical stage ([42] also shows fit with exponential distribution)
$m_1 = 0.01$	rate/person/year for self-referral for diagnosis, from preclinical stage
0.95	specificity of diagnostic test
0.90	sensitivity of diagnostic test for preclinical stage
0.99	sensitivity of diagnostic test for clinical stage
$m_2 = 4.13$	rate/person/year for self-referral for diagnosis, from clinical stage [32]
$r_0 = 100$	treatment completion rate/person/year after a false diagnosis
$r_1 = 0.2$	treatment completion rate/person/year after early diagnosis
$r_2 = 0.2$	treatment completion rate/person/year after late diagnosis
$\delta = 1$	sensitivity of screening for late cancer
$a = 1.5$	service effort for diagnostic test / screening mammogram
$p_1 = 0.05$	probability of recurrence after treatment of an early stage cancer
$p_2 = 0.15$	probability of recurrence after treatment of a late stage cancer

Table 2: Assumed Cost Structure (all in 2003 US$, [47])

Screening mammogram	$145
Diagnostic test	$471

Estimated Treatment Cost per Case of Preclinical and Clinical Stage Breast Cancer

Stage	% at Diagnosis	Est. Discounted Long-term Cost [48]
Preclinical: local	65%	$ 54,013
Clinical: regional	30%	$ 70,066
Clinical: distant	5%	$ 59,463

Table 3: Parameter Values for Numerical Experiments in Section 3.1

Parameter	Values Set for Experiments
Probability of enrollment (h)	0.55, 0.60, 0.65, 0.70, 0.75
Volume standard (std)	480, 2500
Screening service rate (μ_S)	1000, 5000
Number of servers (n)	30, 6

Table 4: Current Situation and Two Improvement Options

Option	Description
0	Current situation: participation=65%, $std = 480$
1	Increase participation: participation=69% $std = 480$
2	Increase minimum accreditation standards: participation=65% $std = 2500$

Table 5: Comparison of Health Outcomes for Current Situation and Improvement Options from Table 4

Option	# Breast Cancer Deaths	# Early Diagnoses	# Late Diagnoses	# Screening Mammograms	# Diagnostic Mammograms	# False Treatments
0	17.0 ± 0.3	25.9 ± 0.25	54.9 ± 0.59	$16,182 \pm 10$	$2,765 \pm 9.84$	133 ± 0.44
1	16.6 ± 0.3	27.4 ± 0.25	53.3 ± 0.59	$17,132 \pm 10$	$2,909 \pm 9.82$	141 ± 0.44
2	16.6 ± 0.16	27.8 ± 0.26	53.5 ± 0.19	$16,176 \pm 12$	$2,107 \pm 5$	101 ± 0.75

Table 6: Cost Summary (US$) for Scenarios in Table 4

Option	Estimated Total Annual Cost of Screening and Treatment
0	$16.0 \times 10^6 \pm 0.1 \times 10^6$
1	$16.6 \times 10^6 \pm 0.1 \times 10^6$
2	$14.0 \times 10^6 \pm 0.14 \times 10^6$

Table 7: Parameters for the Numerical Experiments in Section 3.2

Parameter	Value
Screening interval ($1/f$)	2 years
Number of trained radiologists (n)	4
Max #readings/year/radiologist (μ_S)	2, 500
Initial enrollment probability (h)	(0.35, 0.4, 0.45, 0.5, 0.55, 0.65, 0.75, 0.85, 0.95)
Resource need for diagnostic test (a)	1.5
Target population size	25, 000

Table 8: Simulation Results for Limited Capacity Scenario

h	% Participating	% Screened	Total Demand	# Waiting	Util.	Breast Cancer Deaths
0.35	50	45 ± 0.8	7222 ± 20	2 ± 0.0	0.77	20.3 ± 0.4
0.4	54	53 ± 0.8	7446 ± 18	3 ± 0.1	0.83	20.3 ± 0.4
0.45	57	57 ± 0.8	8263 ± 20	6 ± 0.1	0.88	19.4 ± 0.4
0.5	61	61 ± 0.8	8773 ± 15	13 ± 0.4	0.94	19.4 ± 0.4
0.55	65	64 ± 0.8	9280 ± 10	76 ± 6.7	0.99	19.2 ± 0.4
0.65	73	65 ± 0.7	9450 ± 10	1629 ± 20.6	1.00	19.8 ± 0.4
0.75	81	66 ± 0.9	9530 ± 11	3385 ± 21.5	1.00	20.4 ± 0.5
0.85	88	66 ± 0.8	9600 ± 78	5128 ± 23.5	1.00	21.5 ± 0.4
0.95	96	66 ± 0.9	9670 ± 98	6873 ± 16.6	1.00	22.3 ± 0.5

Table 9: Comparison of Country Statistics with Simulation Results (per 100,000)

Source of Estimate	Incidence	Breast Cancer Deaths
[40] (for 1998)	302.18	70.39
Model Estimate	324	67.8
% Error	7.2%	3.6%

Table 10: American Cancer Society [2] Survival Data

Stage	Pct. at Diagnosis	5-Year Survival Rate	Death Rate
local	65%	96%	0.0081
regional	30%	76%	0.0548
distant	5%	21%	0.312

Table 11: Event Rates

Event			Rate
Ask for screening mammogram	ϕ_1	=	$f \sum_{j=1}^{3} X_{2j} + \gamma \sum_{j=1}^{3} X_{1j}$
Ask for diagnostic mammogram	ϕ_2	=	$m_1 \sum_{i=1}^{2} X_{i2} + m_2 \sum_{i=1}^{3} X_{i3}$
Screening mammogram completion	ϕ_3	=	$\mu_S \sum_{j=1}^{3} X_{4j}$
Diagnostic mammogram completion	ϕ_4	=	$\mu_D \sum_{j=1}^{3} X_{6j}$
Develop preclinical breast cancer	ϕ_5	=	$s_{12} \sum_{i=1}^{6} X_{i1}$
Progress from preclinical to clinical stage	ϕ_6	=	$s_{23} \sum_{i=1}^{6} X_{i2}$
Treatment completion with preclinical stage	ϕ_7	=	$r_1 X_{72}$
Treatment completion with clinical stage	ϕ_8	=	$r_2 X_{73}$
Cancer death	ϕ_9	=	$b_3 \sum_{i=1}^{6} X_{i3} + b_{72} X_{72} + b_{73} X_{73}$
Other death	ϕ_{10}	=	$g \sum_{i=1}^{7} \sum_{j=1}^{3} X_{ij}$

List of Tables

List of Figures